## REMARKS

Claims 24, 25, 29, 30 and 40 are hereby amended to clarify that "text" content is reviewed and processed by the method of the claims. This is not a new issue because it was previously included in claims 27, 31 and 50 and the claims dependent on those claims. This is, however, significant because one of the combining references, Suzuki et al (hereinafter "Suzuki"), lacks any teaching whatever of coding or otherwise dealing with text, its teachings being limited to coding and processing images and color levels, techniques which are not directly transferable to the written (or printed) word.

It is submitted that the Examiner has chosen certain language from Suzuki and applied it to the claims of the present invention in a way which would never be understood as realistic by one skilled in this art and in a way which misrepresents the true meaning of the terms. It is apparent that the operation and significance of Suzuki will have to be explained in somewhat greater detail.

First, it must be recognized that Suzuki is dealing with coding of image signals which have special characteristics, and that its teachings are applicable only to signals which have such special characteristics. As pointed out in Suzuki (col. 1, lines 25 - 33), "image signals have such a tendency that adjacent pixels take similar values, which means that image signals have high correlation. That is, on the frequency axis, the signal power of a high-correlation signal is concentrated on particular frequency components. Therefore, it is possible to reduce the information quantity as a whole by coding coefficients of only those frequency components."

In other words, the Suzuki text is saying that correlated adjacent frequencies tend to be the major components, in terms of amplitude, of the frequency spectrum represented by the image signal and that it is therefore fruitful to select these major frequencies (for the signal interval during which this is so) and record only the changes in amplitude (the coefficients, in Suzuki's wording).

This highlights the first fundamental difference between Suzuki et al and the present invention: the present invention deals with text, and it cannot be said that adjacent components of text information have any such high correlation.

The Examiner may be confusing what might be called "structural compression" with the more familiar "signal compression" approach. The following discussion will attempt to clarify this difference.

When considering image signal compression (as in Suzuki), one is considering the fact that an area of color is defined by the frequencies present and their relative amplitudes and the extent to which the high-amplitude subset of frequencies which define the essential characteristic of the area varies continuously and "smoothly" within the region; an edge is therefore defined by a discontinuity in one or more of the frequencies contributing to the majority of the signal strength up to that point.

This has a superficial similarity to the recognition of a structural information element (such as a word in text) within a stream of information of a particular type, such as a stream of text, as in the present invention.

However, an image signal (or any other signal comprising a composite of frequencies, the mix of which varies over time and

within which characteristic (high-strength) frequencies persist for a significant interval of successive time periods) is divided linearly into its successive elements by correlating its characteristic frequencies at a point in the signal with frequencies at points on either side. The signal points within a region, so defined, are then represented more concisely by reducing the signal to its set of characteristic frequencies and representing the successive values of each frequency comprising the remaining signal by its coefficient relative to the value it took at the previous point in the signal over the region covered, rather than recording its absolute value.

This technique (of identifying and retaining only characteristic frequencies) results in the loss of information from the original signal, such as frequencies beyond the range of human response. This is done extensively in audio signals transmitted across digital carrier networks.

By comparison, the compression of a stream of structured information (such as text) does not proceed according to the production of correlation coefficients between the value of each characteristic frequency found at a point with their values at the points on either side of it, but rather is achieved by reference to information that is external to the point at which the "signal" is being compressed, that is to say, it relies upon "meta" information.

A word of text is identified by the application of meta information that describes the rules of word delimiters, not by the sudden absence of a characteristic frequency at a point on the edge of a region being defined by its presence.

A point in a text stream, unlike a point in a composite frequency signal, has only one scalar dimension. That dimension

is that of representing a single member of the class of the most fundamental structural element defined within the type of structured information being compressed. Within a stream of structured information such as text, a point would be described by the scalar value of a valid member of the fundamental set of characters.

The correlation coefficient of a particular character to adjacent characters in its region, or "word", approaches zero and is therefore not susceptible to the correlation coefficient technique taught by Suzuki in seeking its further compression. Also, since it is a one-dimensional point value, a character is also not susceptible to characterization through selection of lower level attributes, such as frequency elements, as taught by Suzuki.

Froessl teaches that a region of an information stream represented by a set of fundamental points between delimiter points, such as characters comprising a word in a text stream, can be usefully compressed when other identically structured regions can be identified within the context of all past, present and future communications within the network of communication parties under consideration.

In such cases, re-occurring regions (words) may be more concisely represented by shorter codes, the meanings of which are not specific to the compression task performed upon a particular communication, or communication segment, but are made more widely available.

Knowledge of each new code allocated to newly identified regions is extracted from the compression process of every communication compressed, within the context of the communication network under consideration and this meta information is

propagated through the network to nodes requiring to input it into their encoding and decoding activities according to the knowledge acquisition policy of each network node.

Repeated scanning of a document implies, in order to be a useful exercise, the application of suitably adjusted delimiter rules as the process progresses and inevitably results in the widening, in terms of the fundamental elements of the information type under consideration, of the regions being identified. The effect of this is that newly delimited regions begin to subsume groups of regions defined under earlier delimiter rules. When processing a text stream, this results in the allocation of codes to phrases of words.

Thus, the repeated scanning of any text input stream and portions within it and under continually changing delimiter rules results in the identification of larger and larger phrases comprising other phrases and words.

The theoretically optimal compression efficiency can thus be achieved, in the context of any particular communication network, between the benefits of the ultimate size of the compressed output stream and the costs of propagating newly allocated code meanings to other nodes. The careful management of communications and code allocations within and between the contexts of many communication networks enables any communication node to be a member of more than one network and also allows communication networks to be merged or isolated as circumstances dictate in ways which can be transparent to the end users of such a compression approach.

Thus, it would be readily apparent to one skilled in this art that the technique of Suzuki would not be applicable to the problems of the present invention nor to those of Smith et al

with which Suzuki was combined. It therefore cannot be said that combining Suzuki with Smith et al would render the claimed invention obvious.

Secondly, Huffman coding (misspelled "Huffmann" in the Suzuki text) has certain well-known characteristics which preclude a true learning function. As the Examiner is undoubtedly aware, Huffman coding substitutes one or more characters for groups of the original data with the objective that repeated groups of the original data are replaced by a smaller number of characters, resulting in significant compression. The lengths of the codes vary such that the more often repeated sections of data are represented by the shortest code segments and the less frequently repeated sections by longer code segments. Huffman coding has the significant advantage that it yields "lossless" compression, i.e., the result of coding at a transmission end and then decoding at a receiving end yields an exact reproduction of the original data, and becomes more efficient as the length of the message increases. However, it has the disadvantages that, first, the coding scheme in the form of a dictionary must be transmitted as overhead along with the coded message, reducing the efficiency of the process; and, second, that the technique is limited because characters chosen to form the code are necessarily limited in number, meaning that the coding result tends to be rigid and limited in scope.

For coding images, this is not particularly disadvantageous, but for coding text it is a serious disadvantage simply because the variations in the vocabulary of text are vastly greater than the variations in the "vocabulary" of the images being processed.

Efforts have been made to overcome these problems by developing so-called "adaptive" or "dynamic" Huffman coding. The

work of such individuals as Faller, Gallager, Knuth and Vitter, resulted in the FGK and Vitter algorithms. Many discussions of these algorithms are available on the internet and need not be cited here. Basically, the adaptive techniques work by establishing a tree structure in which certain similarities, notably a "sibling" characteristic, are recognized, and then by recalculating the tree as necessary. Since the message sent in coded form includes indications as to when recalculation is needed, both the sender and recipient are aware of the need to do so. In particular, a binary code tree has a sibling property if each node (except the root) has a sibling and if the nodes can be listed in order of non-increasing weight with each node adjacent to its sibling. Both sender and recipient maintain dynamically changing Huffman code trees. The leaves of the code tree represent the source messages and the weights of the leaves represent frequency counts for the messages.

Initially, the code tree consists of a single leaf node, called the 0-node. The 0-node is a special node used to represent the n-k unused messages. For each message transmitted, both parties must increment the corresponding weight and re-compute the code tree to maintain the sibling property. At the point in time when $t$ messages have been transmitted, $k$ of them distinct, and $k<n$, the tree is a "legal" Huffman code tree with $k+1$ leaves, one for each of the $k$ messages and one for the 0-node. If the $(k+1)$st message is one of the $k$ messages already seen, the algorithm transmits $a(t+1)$'s current code, increments the appropriate counter and re-computes the tree. If an unused (i.e., unique) message occurs, the 0-node is split to create a pair of leaves, one for a $(t+1)$, and a sibling which is the new 0-node. Again the tree is re-computed. In this case, the code

for the 0-node is sent; in addition, the receiver must be told which of the $n-k$ unused messages has appeared. At each node a count of occurrences of the corresponding node is stored. Nodes are numbered indicating their position in the sibling property ordering.

From this, it will be seen that the adaptation which occurs with a Huffman code system is a matter of accommodating new messages and, in that very limited sense, constitutes a kind of learning. However, it is not learning of the type described in the claim, wherein, apart from any compression coding which may be used, the system repeatedly and automatically reviews the content of output documents to identify text content that is repeatedly used and which can be replaced by a shorter access code, thereby reducing the volume of unique data to be added to the output documents.

Fundamentally, in Froessl, learning occurs at two levels. First, an encoding node learns a new qualifying word (a region occurring a predetermined number of times in the context of the communication network) and it creates a code to represent it. In so doing, a node learns new words and their associated codes. These codes and words are usually fixed in any particular system.

A word and its code now can be learned by other nodes who might want to encode other occurrences of the word or decode the code when it is received in a communication.

The first stage intra-nodal learning process, in a one-to-many push network, can be as simple as the unilateral allocation of the next available code to each newly "discovered" region (word or phrase). On the other hand, in a many-to-many network, it may involve consultation with other nodes to establish the actual status of a region unknown to itself, and its code, if

already known to others. These process specifics will vary according to its purpose and network circumstances.

The second stage inter-nodal learning process communicates regions and their codes to other nodes. The process specifics will vary according to its purpose and network circumstances.

The essential characteristics of the learning processes being taught by Froessl are therefore not *how* learning is achieved but are *what* is being learned, *why* and *by whom*.

As suggested above, the learning taught by Suzuki operates upon the "lossy" intermediate result of signal point characterization by high energy frequency selection with generation of more or less "lossy" sequential correlation coefficients. On this lossy base, it uses "lossless" Huffman entropy encoding to build a knowledge base of code-region relationships whose purview is limited by the boundaries of the message being compressed. This knowledge is used to communicate these meanings to the decompression process unpacking the specific message at the specific time it is doing so from within the message itself. Knowledge of these meanings has no relevance beyond the specific compressed document in which they are found and the meanings go out of existence with the last copy of that message.

The learning implemented by adaptive Huffman encoding schemes is limited to point-to-point transmission of messages wherein a valid "Huffman tree" is maintained by every node in a communication network for every other node with which it is in communication. Each of these trees is a function of all communications between the two parties to which it relates and the sequence in which these communications occurred. Each copy of a document sent by a party to a set of other parties must be

encoded against the sender's Huffman tree that it holds for compression of messages to that recipient. The efficiency of communication over any one link is a function of the past history of all communications over that one link and every retransmission of any one specific message must also be re-encoded against the then current relevant Huffman tree so as to keep its frequency statistics, and hence its structure, accurate. What is being learned between any two nodes is not the code related to each meaning *per se* but rather is how, for any specific "meaning", its code should be calculated, or recalculated, as the relative occurrence frequency experienced by any one meaning changes.

Froessl teaches that, within any specific communication context, such as a network of nodes, then codes will be permanently allocated to each qualifying "region" (i.e., word or phrase) as it is discovered and this knowledge is communicated as needed anywhere within that context so that it may not only be used by any recipient of a message to decompress it but may also be used by any other node in the compression/decompression of other messages to other nodes in the future, regardless of the continued life or existence of the original message for which the code was created. How this knowledge is learned is not important.

<u>Claim 24</u>

Claim 24 was rejected as being unpatentable over Smith et al (hereinafter Smith) in view of Suzuki. It was again acknowledged in the Action that Smith failed to teach repeatedly and automatically reviewing the content of output documents to identify content that is repeatedly used and which can be replaced by a shorter access code, thereby reducing the volume of

unique data to be added to the output documents. Suzuki was said to teach that function, referring to col. 3, line 66 to col. 4, line 6 and also col. 7, lines 54-67 of Suzuki.

The cited Suzuki patent teaches no such thing.

The first mentioned part of Suzuki teaches that an object of the invention is the improvement of coding efficiency by removing coding redundancy when coding a particular class of documents having image signals and wherein the documents have a large space, a color document having an area with a fixed color, or a document in which a statistical characteristic varies from one area to another. The cited segment also teaches that another object is to increase the processing speed of a decoding process for encoded documents.

The cited portion at col. 7, lines 54-67, is simply an exposition of the well-known entropy encoding technique described above, as an additional and final step to the inventive technique actually taught by Suzuki, which refers to the processing of pixels in a manner which "averages" groups of different pixel values and represents the averaged blocks with symbols. Suzuki then describes using the entropy encoding technique to identify repeating blocks of identical symbols and to allocate shorter or longer other codes to these generated symbols according to their frequency distribution, and finally to store the list of codes and their associated symbol blocks together with the associated substituted string representing the document in encoded form within the generated output document.

At no time or place does Suzuki suggest that repeated and automatic scanning takes place, even on the single document to which the entropy encoding technique is applied, and certainly not across a plurality of output documents at a service center as

is recited in Claim 24.

Furthermore, Suzuki does not suggest that any redundant "document part" generated during his process and then identified by processing be retained for future use as input into subsequent similar processing being performed on subsequent image signal documents belonging to the same client, nor does Suzuki provide for such an eventuality. Thus, Suzuki lacks any teaching of repeatedly and automatically reviewing the text of output documents stored for a client to identify phrases repeatedly used by that client and which are not part of the stored data, and adding the identified phrases to the data comprising parts of the documents to be generated, thereby reducing unique data to be added to output documents.

This is an issue distinct from learning. The purpose of iterative scanning, as claimed, is two fold: it helps identify repetitive regions and it also generates "phrases" comprised of "text words" and "code strings" of unlimited recursive power.

The value of such deep recursion is only realized in a communication context wherein the allocation of codes to delimited regions is permanent such that the knowledge can be propagated as required to any node for purposes of future decoding and unilateral encoding activities.

The adjective "future" is used advisedly. Since codes are permanently assigned, any compressed output document may be stored for immediate future use anywhere in the network so long as its contents have not changed meanwhile. This is not true of adaptive Huffman encoding.

Deep recursion of permanent codes, combined with cross-nodal learning (propagation), also provides for learning optimization whereby a change in a document section represented by a deeply

recursed code in a mature context may only require the propagation of the single high-level code since all codes over which it is defined , even though different from the code it replaces, are all already known to the receiving nodes.

On the other hand, since each node maintains its own set of knowledge bases covering the communication contexts within which it operates, and because the learning processes are independent of the transmission of messages (the compressed versions of which may be stored for *ad hoc* decompression and access as needed do so as to economize on storage space and/or re-compression upon retransmission), it is possible to define contexts within which new meanings may be assigned to stored codes so as to update the content of documents stored in compressed form without reference to the number or location of copies of such documents. Such learning processes may be initiated from any authorized points to participating nodes anywhere within a particular communication context, the efficiency of which processes in terms of network bandwidth will be significantly enhanced by deep recursion within the coding levels of the meanings being replaced.

The process of changing the meaning of a code representing a document section is common to the established techniques in the areas of enterprise document management and team authoring that have developed over the past 25 years. However, the structuring of codes necessary to represent document parts, and latterly also the formatting associated with document sections through customized markup tags, in these contexts is not an autonomous and transparent consequence of the processes, as they can be when they are compressed in a manner informed by the teachings of Froessl.

Accordingly, Claim 24 is allowable because the references,

considered singly or in any proper combination, fail to show or suggest the method of the present invention as recited in claim 24. None of the references of record, alone or in combination, clearly show or fairly suggest repeatedly and automatically reviewing the text content of output documents to identify phrases that are repeatedly used and which can be replaced by a shorter access code, thereby reducing the volume of unique data to be added to the output documents.

Claim 25

In paragraphs 11 and 12 of the Action, Claim 25 was rejected as being unpatentable under §103(a) over Smith in view of Suzuki. As discussed above, Suzuki does not show that for which it was cited either in the specifically identified portions of text or elsewhere in the patent, alone or in any proper combination with other references of record. Accordingly, the rejection is without merit. In addition, Claim 25 further limits claim 24 and is therefore allowable therewith.

Claim 26

Claim 26 was also rejected as being unpatentable over Smith in view of Suzuki. Claim 26 is dependent upon claims 25 and 24 and is allowable therewith. In addition, it should be understood that the document output from a service center contains only the access codes and, usually, unique data. The repeated document parts described by the access codes do not necessarily form part of the document output. This is totally contrary to the teaching of Suzuki wherein the meanings of the access codes, in the form of an associative list of document parts and access codes, together with the document in the form of a list of access codes

and unique data, comprise the main parts of the output document to be transmitted.

Claims 27 and 28

Claims 27 and 28 were also rejected as being unpatentable over Smith and Suzuki. These claims are dependent on claim 24 and should be allowable therewith for the reasons discussed above.

In addition, it should be noted that Suzuki teaches nothing about the text of these claims. The entropy encoding technique used by Suzuki does not comprise the review of output documents to identify phrases repeatedly used but, instead, involves scanning a particular modified image signal in process of becoming a single output document. It cannot, as the Examiner asserts, identify processed image phrases repeatedly used by the particular image signal's owner (the client) either in the subject document or other signal documents, not can it store such phrases anywhere other than the subject document. The rejections of these claims are therefore without basis.

Claim 29

Claim 29 is independent but also includes language which clearly distinguishes the claim from the teachings of Smith and/or Suzuki. Specifically, claim 29 recites reviewing automatically and in a learning mode the content of the output documents to identify parts thereof that are repeatedly used amongst such documents, generating automatically a storage access code uniquely associated with such identified document parts and adding the identified document parts, each with its uniquely associated storage access code, to the stored data comprising

parts of documents to be compiled, compiling and storing output documents of selected format and content and designated unique data by substituting in response to requests from clients the storage access codes of the document parts identified in and by document output compilation requests from clients, and transmitting the output documents to a recipient service center, thereby reducing unique data to be added to output documents.

Neither Smith nor Suzuki includes a teaching of a learning mode meeting the requirements of this language. Accordingly, claim 29 should be allowed.

<u>Claims 30 and 31</u>

Claims 30 and 31 include language such as "repeatedly and automatically reviewing the text of output documents stored for a client to identify phrases repeatedly used by that client and which are not part of the stored data, and adding the identified phrases to the data comprising parts of documents to be generated, thereby reducing unique data to be added to output documents". This language is in combination with other features of the system. As pointed out above, a teaching of this aspect of the invention is nowhere taught by Suzuki or Smith, alone or in any proper combination, and claims 30 and 31 should therefore be allowable.

<u>Claims 32-48</u>

Claims 32 through 48 are all ultimately dependent on claim 31 and should therefore be allowable therewith. While these claims may also include other distinguishing features, they are allowable whether or not that is true and it does not appear to be necessary to discuss such features separately.

<u>Claims 50 and 51</u>

Claims 50 and 51 were rejected as being anticipated by Smith. However, claim 50, among other features, recites repeatedly and automatically reviewing the text of output documents stored for a client to identify phrases repeatedly used by that client and which are not part of the stored data, and adding the identified phrases to the data comprising parts of documents to be generated, thereby reducing unique data to be added to output documents. It was acknowledged in the Action (with reference to other claims) that this feature is not shown by Smith. Furthermore, as pointed out above, this feature is not taught by Suzuki. Accordingly, the rejection on this basis of claim 50 is clearly in error and should be withdrawn.

Claim 51 is dependent on claim 50 and should be allowable therewith.

<u>Claims 52 and 53</u>

Claims 52 and 53 were separately rejected as being unpatentable over Smith in view of Perry which was cited as showing means for storing and printing documents for sale. However, claims 52 and 53 are dependent on claim 50 which, as described above is allowable as including repeatedly and automatically reviewing the text of output documents stored for a client to identify phrases repeatedly used by that client and which are not part of the stored data, and adding the identified phrases to the data comprising parts of documents to be generated, thereby reducing unique data to be added to output documents. Perry does not show any such feature, nor does any other reference cited and relied upon in the Action. Since

claims 52 and 53 are dependent on claim 50, they should be allowable therewith, regardless of whatever storing and printing features are shown in Perry.

Counsel apologizes for the length of the discussions herein. However it is felt that a rather complete understanding of the technology of the prior art is essential to an understanding of the contributions of the claimed invention and it is hoped that the discussions have provided that understanding.

Respectfully submitted,

Walter C. Farley
Reg. No. 22,624
Attorney for Applicant

P.O. Box 329
Harpswell ME 04079
Fax 207 833 7831
Tel 207 833 6793